第47卷 第3期
2024年3月

干旱区地理
ARID LAND GEOGRAPHY

Vol. 47 No. 3
Mar. 2024

组合光学和微波遥感的耕地土壤含盐量反演

刘瑞亮[1]， 贾科利[1]， 李小雨[1]， 陈睿华[1]， 王怡婧[1]， 张俊华[2]

（1. 宁夏大学地理科学与规划学院，宁夏 银川 750021；2. 宁夏大学生态学院，宁夏 银川 750021）

摘 要：耕地保护关系到国家粮食安全和经济社会可持续发展，对生态环境保护具有重要作用，快速精准的获取耕地土壤盐分含量及空间分布信息是耕地保护的必然要求。以宁夏平罗县为研究区，利用Landsat 9 OLI和Sentinel-1遥感影像，提取光谱指数和雷达极化组合指数，基于变量投影重要性法与灰度关联法筛选特征变量，然后运用反向传播神经网络、支持向量机和随机森林3种机器学习算法构建模型，并用最佳模型反演耕地土壤含盐量空间分布情况。结果表明：(1) 利用变量投影重要性法筛选变量建立的模型验证集决定系数(R^2)大于灰度关联法筛选变量建立的模型。(2) 利用随机森林算法，组合光谱指数和雷达极化组合指数协同反演模型效果最佳，建模集R^2为0.791，均方根误差(RMSE)为1.016，R^2较单一数据源模型分别提高0.065和0.085，RMSE分别降低0.147和0.189；验证集R^2为0.780，RMSE为1.132，R^2较单一数据源模型分别提高0.091和0.237，RMSE分别降低0.175和0.377。(3) 平罗县耕地轻度盐渍化和中度盐渍化土壤分布范围广，占比分别为23.77%和33.54%，重度盐渍化达15.37%。研究结果发现，组合多源遥感数据建模能够有效提高土壤含盐量反演精度，可为干旱区耕地土壤含盐量的反演和当地农业可持续发展提供有效的技术参考。

关 键 词：光学和微波遥感；机器学习；耕地；土壤含盐量反演

文章编号：1000-6060(2024)03-0433-12(0433~0444)

自然或人为引起的土壤盐渍化是一种重大的环境灾害，广泛发生在内陆干旱、半干旱地区，严重影响农业生产和区域可持续发展。据估计，到2050年全球约50%的耕地会面临盐碱化[1]，将成为世界性的生态问题[2]。因此，运用科学手段精准反演和监测耕地土壤盐渍化动态，及时治理和缓解耕地土壤盐渍化，对区域粮食生产和农业可持续发展具有重要意义。

卫星遥感手段能快速、宏观地获取土壤光谱特征，通过构建遥感监测模型，可以实现大范围土壤盐渍化监测和评估[3]。光学遥感可以提供多波段的光谱信息，对土壤盐分具有较强敏感性，研究者通过构建光谱指数实现土壤盐分的反演。陈红艳等[4]基于Landsat 8 OLI，引入第7波段(2100~2300 nm)对植被指数进行改进，模型反演精度大幅提升。Wang等[5]等用Sentinel-2红边波段代替原来红光波段构建光谱指数，提高了光谱信息对土壤盐分的敏感性。然而单一光学数据易受成像时间、云雨等天气影响，依赖光谱特征提取盐渍化信息具有局限性[6]。微波遥感具有全天候工作能力，不易受气象条件和日照水平影响。同时，由于试验区有植被覆盖，光谱反射率受影响，而微波遥感能穿透植被，具有探测地表下目标的能力[7]。马驰等[8]基于Sentinel-1雷达影像，分析垂直极化模式(V_V)和交叉极化模式(V_H)组合的后向散射系数与土壤盐分之间的关系，确定$(V_V^2+V_H^2)/(V_V^2-V_H^2)$极化组合的后向散射系数可以较好分离不同含盐量的土壤。张智韬等[9]利用Sentinel-1雷达数据探究了内蒙古河套

收稿日期：2023-07-21； **修订日期**：2023-08-25

基金项目：国家自然科学基金项目(42061047，42067003)；国家重点研发计划项目(2021YFD1900602)；宁夏回族自治区重点研发计划项目(2021BEG03002)资助

作者简介：刘瑞亮(1998-)，女，硕士研究生，主要从事遥感监测与分析研究. E-mail: liuruiliang2022@163.com

通讯作者：贾科利(1975-)，男，博士，教授，主要从事3S与土地利用. E-mail: jiakl@nxu.edu.cn

灌区不同深度土壤含盐量。尽管上述研究已经取得较好的结果,但多基于单一遥感数据,而组合光学和微波遥感反演土壤含盐量相关研究较少。肖森天等[10]组合光学影像和雷达后向散射特征,验证了光学与雷达数据组合在土壤盐渍化监测方面的可行性。然而该研究仅利用随机森林分类算法对土壤盐渍化进行了等级分类,并没有反演出土壤不同区域具体的盐分值。而本文利用随机森林的回归算法,能够反演得到研究区耕地像元尺度土壤含盐量,在一定程度上可以保留更多的盐分信息。

土壤盐渍化产生原因和土壤盐分组成复杂,在特征变量的选择上,不同区域的研究结果存在差异[11]。因此,需要采用不同方法来筛选特征变量。常用的方法有相关系数法(PCC)、逐步回归(SR)、灰度关联法(GC)和变量投影重要性法(VIP)等。王海峰[12]采用GC、SR和VIP 3种方法筛选变量构建模型,发现不同筛选方法下模型精度不同。Wang等[5]基于PCC、VIP、GC、随机森林(RF)筛选变量,并利用偏最小二乘回归(PLSR)建立模型,结果显示GC-PLSR模型决定系数(R^2)高于VIP-PLSR模型。不同建模方法反演效果不同,刘恩等[13]等采用多元线性回归和反向传播神经网络(BPNN)反演小开河引黄灌区土壤盐渍化,发现BPNN的精度优于传统的多元线性回归。而陈红艳等[4]以Landsat 8 OLI建立的土壤含盐量反演模型中,支持向量机(SVM)精度优于BPNN。此外,RF具有可处理非线性数据、抗拟合能力强的特点,在土壤盐分及主要离子定量估测方面已显示出较好效果[14]。变量筛选方法结合机器学习算法进行土壤盐分反演是近几年的研究热点,而比较不同变量筛选方法,并结合不同机器学习算法反演土壤含盐量相关研究较少。

本文以宁夏平罗县为研究区,基于光学和微波遥感影像,提取光谱指数和雷达极化组合指数,使用VIP和GC筛选特征变量,然后利用BPNN、SVM和RF机器学习算法构建耕地土壤含盐量反演模型,评定不同变量输入和不同建模方法下土壤含盐量反演精度,并反演研究区耕地土壤盐分分布,为银川平原耕地土壤盐渍化的潜在识别和防治提供理论依据和技术支持。

1 数据与方法

1.1 研究区概况

研究区位于宁夏石嘴山市平罗县(38°26′60″~39°14′09″N,105°57′40″~106°52′52″E),地处贺兰山东麓洪积扇和黄河冲积之间、宁夏平原北部,是重要的灌溉农业区。年平均降水量150~203 mm,年蒸发量超过1825 mm,蒸降比达到10:1。主要土地利用类型为耕地、林地、草地和盐碱荒地,耕地总面积约9.5×10^4 hm^2,主要分为水田和旱地。试验区(图1)

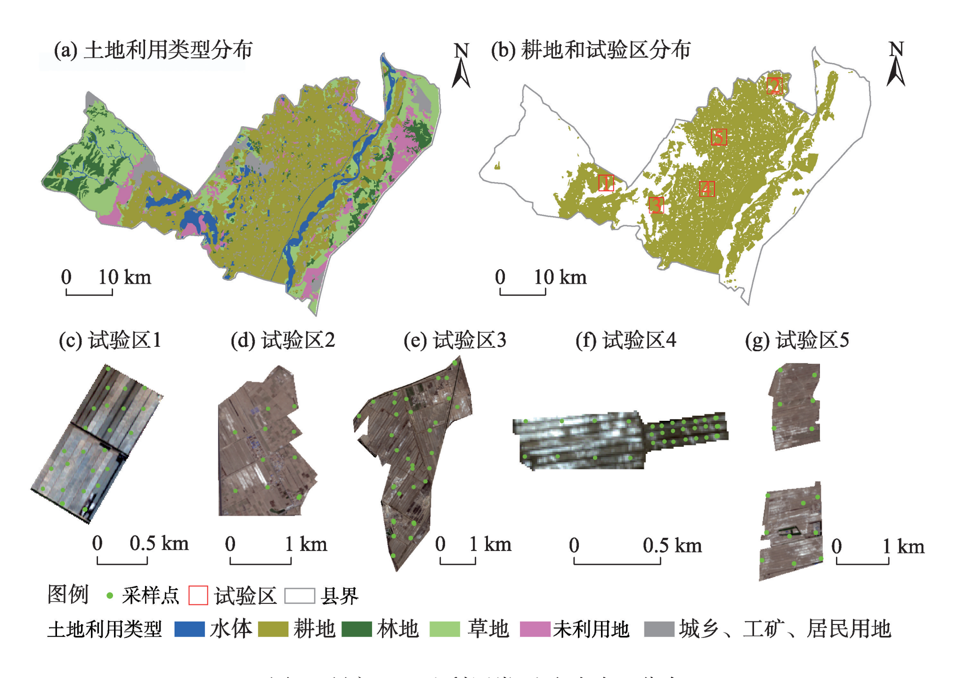


图1 研究区土地利用类型及试验区分布

Fig. 1 Land use types of the study area and distribution of the test areas

选择平罗县耕地区域进行布设，分别位于平罗县山前洪积扇区、西大滩碟形洼地、冲积平原区、灵盐台地和河滩区五大地貌单元。同时，采样时间处于春耕农忙之前，人为扰动较少。

1.2 数据来源与预处理

1.2.1 土壤样本采集与盐分测定 样本采集于2022年3月20日—4月8日，采样时利用五点法采集0~20 cm土壤，编号装入采样袋带回实验室。土样经过自然风干、研磨，并用2 mm孔径筛过滤后，以水土5:1配置提取液[15]，运用电导率法[16]计算样本的含盐量，每个样本重复3次，取平均值作为该点的盐分值。

按含盐量大小划分样本盐渍化程度[17]。样本总体变异系数为79.154%（表1），呈中等强度变异，表明样本分布离散，具有普适性。将样本按含盐量大小进行排序，每隔2个样本取出1个用于模型验证，剩余样本用于建模，共选取70个样本作为建模数据，34个样本作为验证数据。建模集与验证集样本分布趋势与总样本分布趋势一致（图2），说明样本划分合理，可用于模型构建和验证。

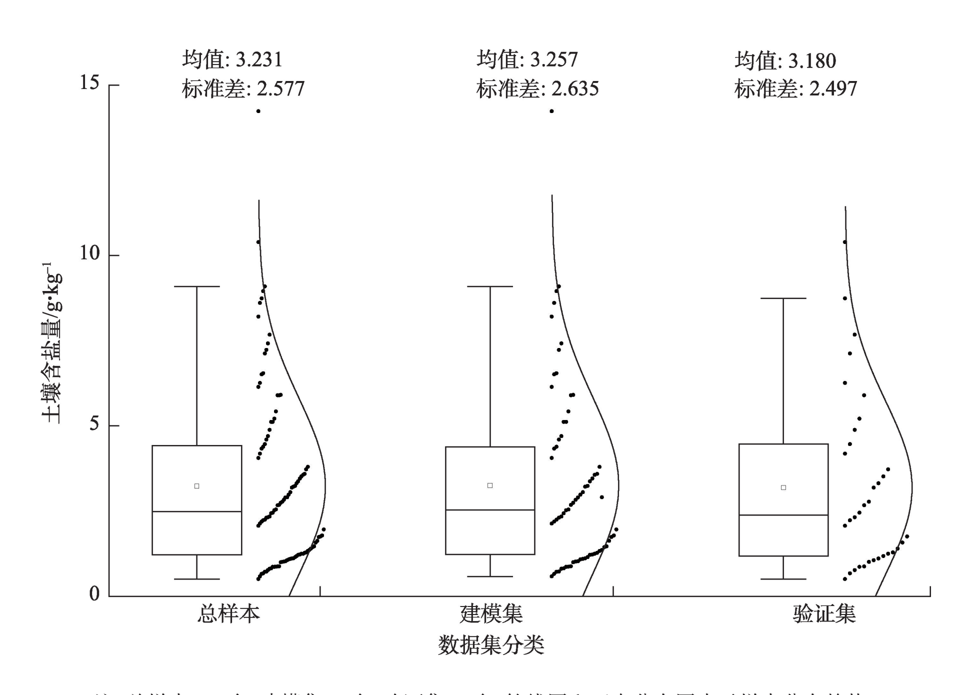
1.2.2 遥感影像数据获取与处理 Landsat 9 OLI遥感影像来源于美国地质勘探局（https://earthexplorer.usgs.gov），行列号为129/33，成像时间为2022年4月5日，云量为4.7%。在ENVI 5.3中对影像进行辐射校正、裁剪等预处理，计算盐分指数和植被指数，如表2所示。

Sentinel-1数据来源于欧洲航空航天局数据网站（https://scihub.copernicus.eu），成像时间为2022年

表1 土壤样本描述性统计

Tab. 1 Descriptive statistics of soil samples

样本等级(含盐量/g·kg⁻¹)	样本数量/个	含盐量/g·kg⁻¹			变异系数/%
		平均值	最大值	最小值	
非盐渍化(<1)	15	0.775	0.999	0.510	16.712
轻度盐渍化(1~2)	27	1.326	1.963	1.008	19.913
中度盐渍化(2~4)	32	2.853	3.795	2.066	18.081
重度盐渍化(4~6)	15	4.966	5.909	4.060	12.385
盐土(>6)	15	8.208	14.231	6.145	25.133
总样本	104	3.251	14.231	0.510	79.154



注：总样本104个，建模集70个，验证集34个；箱线图和正态分布图表示样本分布趋势。

图2 不同数据集土壤样本分布

Fig. 2 Distributions of soil samples in different datasets

表2 光谱指数计算公式

Tab. 2 Calculation formulas of spectral indices

光谱指数	公式
盐分指数(SI)[18]	$\sqrt{Blue \times Red}$
盐分指数1(SI1)[19]	$\sqrt{Green \times Red}$
盐渍化指数1(S1)[20]	Blue/Red
盐渍化指数2(S2)[21]	(Blue−Red)/(Blue+Red)
盐渍化指数5(S5)[22]	Blue×Red/Green
扩展差值植被指数(EDVI)[4]	NIR+SWIR1−Red
扩展比值植被指数(ERVI)[4]	(NIR+SWIR1)/Green
大气阻抗植被指数(ARVI)[23]	(NIR−2Red+Blue)/(NIR+2Red−Blue)
冠层盐度响应植被指数(CRSI)[24]	$\sqrt{\frac{(NIR \times Red)-(Green \times Blue)}{(NIR \times Red)+(Green \times Blue)}}$
扩展增强型植被指数(EEVI)[4]	$\frac{2.5(NIR+SWIR1-Red)}{(NIR+SWIR1+6Red-7.5Blue+1)}$

注:Blue、Green、Red、NIR、SWIR1分别为蓝波段(b2)、绿波段(b3)、红波段(b4)、近红外波段(b5)、短波红外1(b6)的光谱反射率(%)。

表3 雷达极化组合指数计算公式

Tab. 3 Calculation formulas of radar polarization combination indices

雷达极化组合指数	参考文献	雷达极化组合指数	参考文献
V_V	[9]	$V_V^2-V_H^2$	[9]
V_H	[9]	$V_V^2+V_H$	[9]
V_V-V_H	[8]	$V_V^2-V_H$	[9]
V_V+V_H	[8]	$V_H^2-V_V$	[9]
$V_V^2+V_H^2$	[9]	$V_H^2-V_V^2$	[9]

注:V_V为垂直极化的雷达后向散射系数;V_H为交叉极化的雷达后向散射系数。

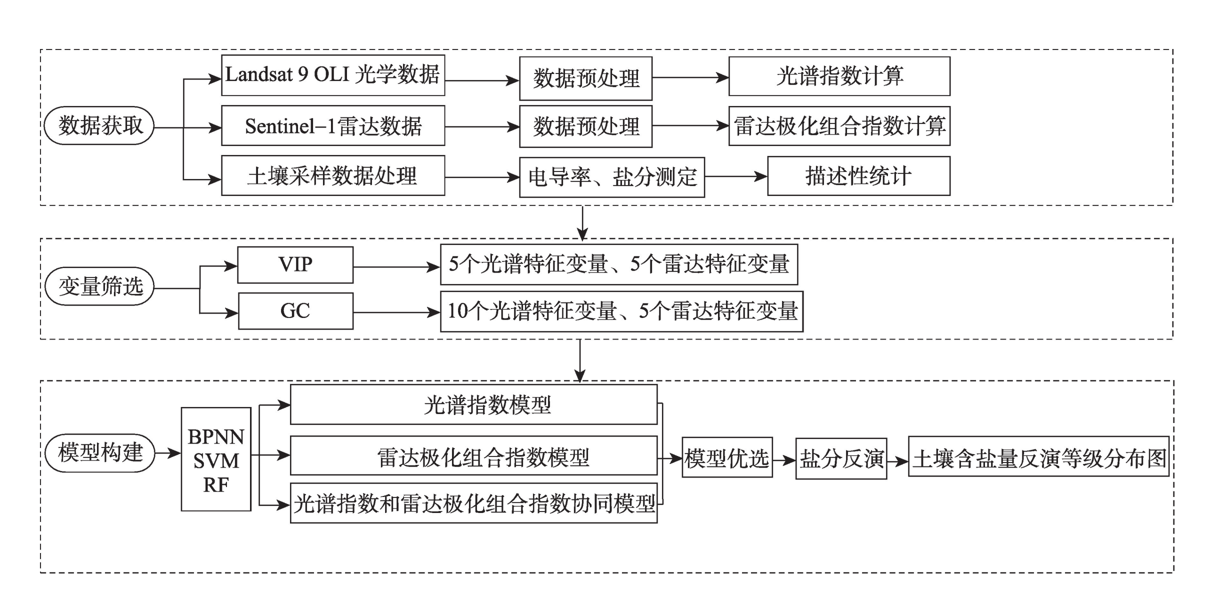
4月1日,数据级别为Level 1,包括垂直极化模式和交叉极化模式。利用SNAP软件对数据进行热噪声去除、轨道文件校正、辐射定标等处理。已有研究指出[25],对于单极化雷达数据,如提取的土壤信息量相对较少,研究结果会受到一定影响。因此,将雷达影像的极化方式进行组合,以提高土壤含盐量反演精度[26-27]。本文所用指数如表3所示。

1.2.3 土地利用数据获取及处理 土地利用数据从中国科学院资源环境科学与数据中心(http://www.resdc.cn)获取,根据分类标准将研究区土地利用分为6类(图1):耕地、林地、草地、水域、未利用地以及城乡、工矿、居民用地,提取耕地作为研究对象。

1.3 研究方法

首先进行数据获取和处理,包括光谱指数、雷达极化组合指数的计算,土壤含盐量的测定;然后采用变量投影重要性法和灰度关联法进行土壤含盐量特征变量筛选;最后利用筛选的变量,采用3种不同机器学习算法分别构建光谱指数模型、雷达极化组合指数模型、光谱指数和雷达极化组合指数协同模型,并进行精度验证,选择最优模型进行研究区耕地土壤含盐量反演。具体研究思路如图3所示。

1.3.1 土壤含盐量特征变量筛选 采用变量投影重要性法和灰度关联法进行特征变量筛选。变量投



注:VIP、GC分别为变量投影重要性法和灰度关联法;BPNN、SVM、RF分别为反向传播神经网络、支持向量机和随机森林。

图3 研究路线

Fig. 3 Research route

影重要性法是一种基于偏最小二乘回归的变量筛选方法[28]。对于给定的自变量，变量投影重要性不仅表示自变量对因变量的影响，还考虑了其他自变量对因变量的间接影响。计算公式为：

$$VIP_j = \sqrt{\frac{p \times \sum_{f=1}^{F} SSY_f \times W_{jf}^2}{SSY_{total} \times F}} \tag{1}$$

式中：VIP_j 为 j 变量的重要性；p 为自变量个数；F 为主成分总数；f 为主成分；SSY_f 为 f 主成分解释的方差平方和；SSY_{total} 为因变量平方和；W_{jf}^2 为 j 变量在 f 主成分中的重要性。VIP_j 越大，自变量对因变量的解释力越强。当自变量的 VIP_j 大于1时，独立变量被判断为重要的自变量[29]。

灰色关联度分析是一种多因素统计方法，其目的是通过一定的方法原理确定系统中各因素的主要关系，用灰色关联度刻画因素间关系的强弱和次序，找出影响最大的因素[30]。计算公式为：

$$GCD = \frac{1}{n}\sum_{t=1}^{n} \gamma[x_0(t), x_i(t)] \tag{2}$$

式中：GCD为灰色关联度；t 为变量；n 为变量个数；γ 为关联系数；$x_0(t)$ 为参考序列；$x_i(t)$ 为比较序列，其中：

$$\gamma[x_0(t), x_i(t)] = \frac{\min_i \min_t |x_0(t) - x_i(t)| + \rho \max_i \max_t |x_0(t) - x_i(t)|}{|x_0(t) - x_i(t)| + \rho \max_i \max_t |x_0(t) - x_i(t)|} \tag{3}$$

式中：$\min_i \min_t |x_0(t) - x_i(t)|$ 和 $\max_i \max_t |x_0(t) - x_i(t)|$ 为极差最小值和最大值；ρ 为介于[0, 1]之间的分辨系数，本文设为0.5。

1.3.2 机器学习模型

（1）反向传播神经网络（BPNN）

BPNN基于反向传播误差方式对数据集进行训练，达到误差最小化的目的[31]。该模型是对非线性映射的全局逼近，拥有较强的自适应和自学习能力[32]。本文模型训练目标最小误差设置为0.00001，迭代次数设置为1000次，学习率设置为0.01。

（2）支持向量机（SVM）

SVM基于结构风险最小原理，依靠有限的样本来检索全局最优解，对未知点有较好的泛化效果[33]。本文选用RBF（Radial basis function）为SVM核函数类型，惩罚因子（c）与核函数（g）由样本训练得到，分别为13和5。

（3）随机森林（RF）

RF是基于多棵回归树的集成学习方法，通过结合多个决策树，并平均其结果使得决策树泛化误差收敛从而产生更好的预测结果[34]。RF善于处理变量间的非线性关系，预测性能受回归树棵数、最大深度、最小叶子数等参数影响[35]。本文回归树棵树为200，最小叶子数为4。

1.3.3 模型评价 为定量比较不同算法反演土壤含盐量的精度，本文选择2个常用指标决定系数（R^2）和均方根误差（RMSE）来进行量化。R^2越大，RMSE越小，说明模型拟合效果越好。

2 结果与分析

2.1 土壤含盐量的特征变量筛选

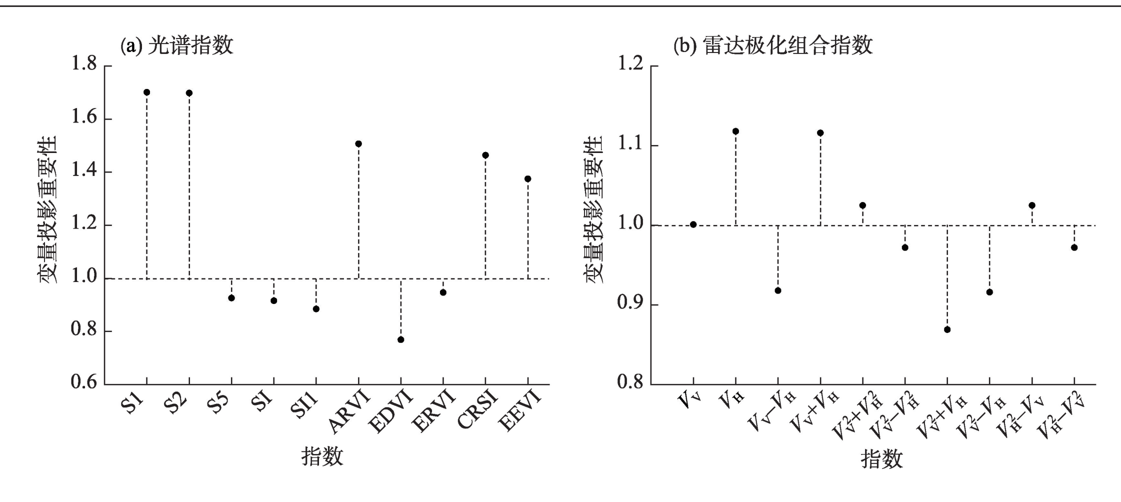
利用式（1）、式（2）计算指数的变量投影重要性（图4）和灰色关联度（图5），筛选出特征变量参与建模。结果表明，光谱指数S1、S2、ARVI、CRSI、EEVI（图4a）和雷达极化组合指数 V_V、V_H、V_V+V_H、$V_V^2+V_H^2$、$V_H^2-V_V$（图4b）重要性大于1，说明这些变量是指示土壤盐分的重要变量，可作为特征变量参与建模。

由图5a可知，光谱指数与土壤含盐量灰色关联度均大于0.73，其中ARVI灰色关联度最高，达到了0.784；S2灰色关联度最低，为0.738。雷达极化组合指数与土壤含盐量灰色关联度（图5b）介于0.71~0.75，最高为 V_V+V_H 的0.739。为实现特征变量筛选，本文设光谱指数和雷达极化组合指数灰色关联度阈值为0.720，统计基于灰度关联分析的土壤含盐量敏感指数情况。结果发现，利用灰度关联法筛选的光谱特征变量较雷达特征变量多，其中，10个光谱指数都为特征变量，雷达极化组合指数中筛选出 V_V、V_H、V_V-V_H、V_V+V_H 和 $V_V^2-V_H$ 共计5个特征变量参与建模。

2.2 模型构建与验证

光谱指数模型中，以土壤含盐量作为因变量，选取VIP筛选的S1、S2、ARVI、CRSI、EEVI作为自变量构建VIP-BPNN、VIP-SVM、VIP-RF模型；选取GC筛选的S1、S2、S5、SI、SI1、ARVI、EDVI、ERVI、CRSI、EEVI作为自变量构建GC-BPNN、GC-SVM、GC-RF模型。

雷达极化组合指数模型中，选取VIP筛选的 V_V、V_H、V_V+V_H、$V_V^2+V_H^2$、$V_H^2-V_V$ 作为自变量构建VIP-BPNN、VIP-SVM、VIP-RF模型；选取GC筛选的 V_V、V_H、



注：横轴光谱指数和雷达极化组合指数含义同表2、表3所示。下同。

图4 特征变量与土壤含盐量的投影重要性

Fig. 4 Projection importance between soil salt content and characteristic variables

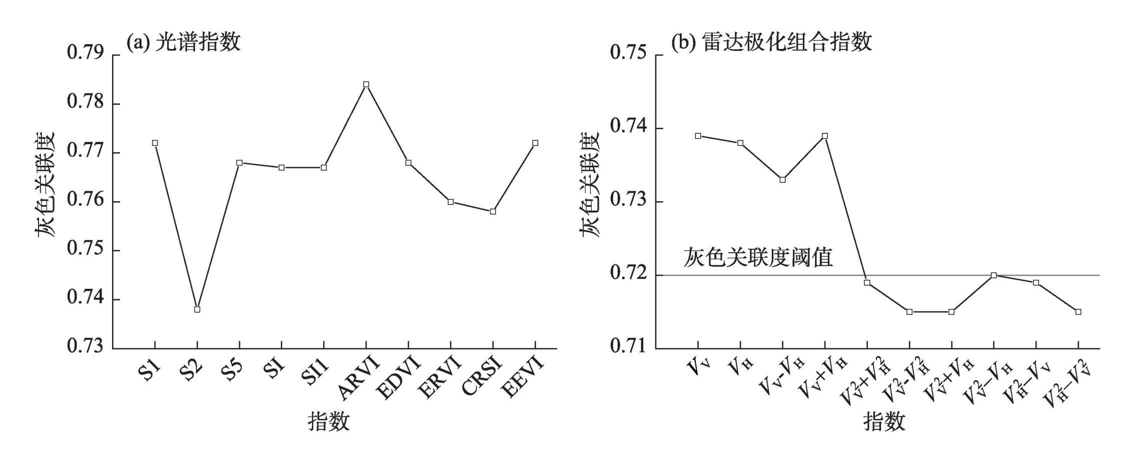


图5 特征变量与土壤含盐量的灰色关联度

Fig. 5 Grey relativity between soil salt content and characteristic variables

V_V-V_H、V_V+V_H、$V_V^2-V_H$作为自变量构建GC-BPNN、GC-SVM、GC-RF模型。

结果发现(表4)，光谱指数模型中，利用VIP筛选变量构建的3个模型验证集R^2均大于0.46，RMSE均小于1.93，其中VIP-RF模型精度最高，建模集和验证集R^2分别为0.726和0.689，RMSE分别为1.163和1.307，模型具有较强的稳定性。利用GC筛选变量构建的3个模型验证集R^2整体较低，其中GC-SVM模型验证效果较好，GC-BPNN模型次之，GC-RF模型验证集误差最大。雷达极化组合指数模型中(表4)，VIP筛选变量建立的模型建模集和验证集R^2呈现VIP-RF>VIP-SVM>VIP-BPNN的规律；GC筛选变量建立的模型建模集和验证集R^2呈现同样的规律，表明随机森林机器学习算法在雷达极化组合指数模型中具有较好的反演能力。比较光谱指数模型和雷达极化组合指数模型发现，光谱指数模型精度普遍高于雷达极化组合指数模型，这可能与本文盐分采样为0~20 cm表层土壤有关，雷达数据更适合于探测地表下目标，而光学数据具有丰富的地表反射率信息。

对比2种变量筛选方法建立的模型精度(表4)发现，光谱指数模型中，验证集精度VIP-RF>GC-RF，VIP-BPNN>GC-BPNN，VIP-SVM>GC-SVM。雷达极化组合指数模型中，除VIP-BPNN<GC-BPNN模型外，其他2种机器学习模型也呈现出基于VIP方法筛选变量构建模型精度大于基于GC筛选变量构建模型精度的规律，表明VIP方法筛选的变量较优，更适于研究区土壤含盐量反演。因此，为验证多源数据对土壤含盐量反演精度的影响，将经过VIP方法筛选得到的光谱指数S1、S2、ARVI、CRSI、EEVI和

表4 基于单一遥感数据的机器学习模型

Tab. 4 Machine learning models based on single remote sensing data

变量	模型类别	建模集		验证集	
		R^2	RMSE	R^2	RMSE
光谱指数	VIP-BPNN	0.325	1.828	0.464	1.926
	VIP-SVM	0.552	1.487	0.476	1.840
	VIP-RF	0.726	1.163	0.689	1.307
	GC-BPNN	0.328	1.822	0.433	1.979
	GC-SVM	0.488	1.591	0.461	1.904
	GC-RF	0.649	1.317	0.327	2.158
雷达极化组合指数	VIP-BPNN	0.228	1.954	0.291	2.214
	VIP-SVM	0.482	1.475	0.524	1.814
	VIP-RF	0.706	1.205	0.543	1.509
	GC-BPNN	0.220	1.964	0.346	2.127
	GC-SVM	0.509	1.440	0.356	1.910
	GC-RF	0.518	1.543	0.413	2.015

注：VIP-BPNN为基于变量投影重要性法筛选变量建立的反向传播神经网络模型；VIP-SVM为基于变量投影重要性法筛选变量建立的支持向量机模型；VIP-RF为基于变量投影重要性法筛选变量建立的随机森林模型；GC-BPNN为基于灰度关联法筛选变量建立的反向传播神经网络模型；GC-SVM为基于灰度关联法筛选变量建立的支持向量机模型；GC-RF为基于灰度关联法筛选变量建立的随机森林模型；R^2为决定系数；RMSE为均方根误差。下同。

雷达极化组合指数V_V、V_H、V_V+V_H、$V_V^2+V_H^2$、$V_H^2-V_V$共同作为自变量参与建模，结果如表5。

组合光谱指数和雷达极化组合指数模型反演精度较单一光谱指数和雷达极化组合指数模型大幅提高。其中，VIP-BPNN模型验证集R^2分别提高0.124和0.297，RMSE分别降低0.237和0.525。VIP-SVM和VIP-RF模型建模集和验证集R^2均大于0.6，RMSE均小于1.6，模型准确度和学习性能更强。其中，VIP-RF建模集R^2较单一光谱指数和雷达极化组合指数模型分别提高0.065和0.085，R^2达到了0.791，RMSE分别降低0.147和0.189；验证集R^2分别提高0.091和0.237，R^2达到了0.780，RMSE分别降低0.175和0.377，表明光谱指数和雷达极化组合指数协同反演模型效果优于单一数据源模型，说明多源遥感数据参与建模能够有效提高土壤含盐量预测精度。

图6为基于VIP-RF构建的光谱指数模型，雷达极化组合指数模型以及协同模型验证集的实测值与预测值的散点图。结果表明，组合光谱指数和雷达极化组合指数共同建立的VIP-RF模型拟合效果最好，故选择此模型对研究区耕地土壤含盐量进行反演。

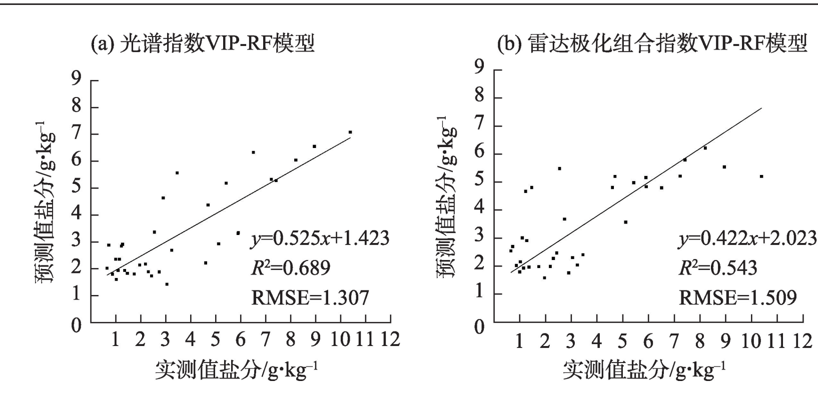
2.3 模型应用

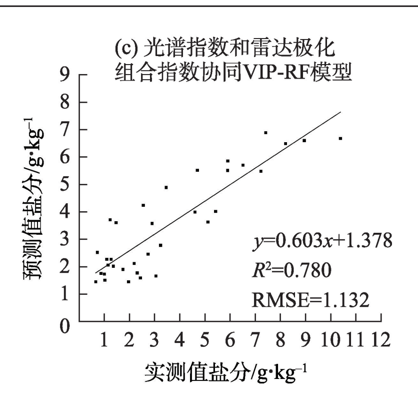
选择精度最佳的光谱指数和雷达极化组合指数VIP-RF协同反演模型对平罗县耕地进行土壤盐分反演，得到平罗县耕地土壤盐分反演等级图（图7），利用ArcGIS统计不同盐渍化等级土壤面积（表6）。由结果可知，平罗县耕地盐渍化情况较为严重，盐渍化土壤占比达到了耕地总面积的86.81%，中度盐渍化高达33.54%，重度盐渍化和盐土主要分布在中西部和东部引黄灌溉区。盐渍化反演结果与实地采样情况较为一致，表明组合多源遥感数据构建反演模型在研究区耕地土壤盐分反演中具有可行性。

表5 基于多源遥感数据的机器学习模型

Tab. 5 Machine learning models based on multi-source remote sensing data

变量	模型类别	建模集		验证集	
		R^2	RMSE	R^2	RMSE
光谱指数和雷达极化组合指数	VIP-BPNN	0.694	1.230	0.588	1.689
	VIP-SVM	0.787	1.027	0.643	1.572
	VIP-RF	0.791	1.016	0.780	1.132





注：R^2 为决定系数；RMSE为均方根误差。

图6 不同数据源变量的VIP-RF模型

Fig. 6 VIP-RF models of different data source variables

表6 耕地土壤含盐量反演等级统计

Tab. 6 Grade statistics of soil salinity inversion in cultivated land

盐渍化程度	非盐渍化	轻度盐渍化	中度盐渍化	重度盐渍化	盐土
像元数/个	138897	250353	353408	161882	148863
面积/km^2	124.95	225.18	317.73	145.60	133.86
占比%	13.19	23.77	33.54	15.37	14.13

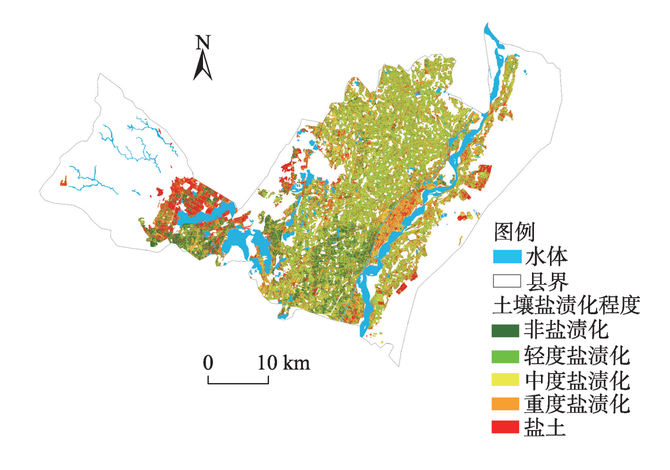


图7 耕地土壤含盐量反演等级分布

Fig. 7 Grade distribution of soil salinity inversion in cultivated land

3 讨 论

组合多源遥感数据反演土壤含盐量已成为近年来研究热点，但如何高效组合不同数据源，减少影像数据处理过程中存在的部分光谱特征、雷达信息丢失和失真等问题，还有待继续研究。余祥伟等[36]利用乘积变换、G-S变换、PC变换、小波变换等算法将星载SAR和Landsat 8 OLI进行融合，发现G-S变换和PC变换能够较好实现星载SAR和光学影像的融合，但仍然存在信息量、空间分辨率增加的同时，光谱特征丢失的问题。姜红等[37]研究发现，以雷达后向散射系数和改进型温度植被干旱指数共同作为SVM模型输入参数时，土壤水分监测精度显著提高。因此，本文组合VIP筛选得到的光谱指数和雷达极化组合指数共同参与建模，充分结合光学遥感和微波遥感优势，较好地保留了土壤的光谱特征和雷达信息。

最优变量筛选方法结合最优机器学习算法可以有效提高反演精度。本文比较VIP和GC筛选变量优劣，发现VIP的结果更为可靠，与王海峰等[12]的研究结果相同。这是因为VIP通过判断特征对模型准确率的影响和加入噪声干扰前后模型准确率的变换来实现对模型精度的优化构建，但李明亮[38]指出GC筛选变量建模同样能够有效提高模型拟合精度。有研究[39]指出，土壤成分对反演模型的性能有很大的影响，因此本文所选变量筛选方法是否适用于其他研究区有待商榷。本文选择BPNN、SVM和RF 3种传统机器学习算法反演土壤盐分，发现BPNN和SVM精度较低，RF精度最高，能够满足研究区土壤含盐量的反演，这与张智韬等[40]所得结论基本一致。可见，将RF应用于土壤含盐量反演，能够在一定程度上提升模型反演精度。而近年来，诸多学者尝试引入网格搜索、随机搜索和贝叶斯等算

法对机器学习进行超参数优化，Chen等[41]提出了一种基于树结构Parzen估计器（TPE）优化算法的极端梯度提升（XGBoost）模型，结果表明TPE联合优化算法显著提高了XGBoost模型的性能，具有较强的泛化能力。Wang等[42]同样在贝叶斯优化框架下，将超参数和特征选择相结合，对光梯度增强机（LightGBM）模型进行自适应联合优化。Xu等[43]提出了一种自适应遗传算法（AGA）的支持向量机输入特征和超参数同时识别的新方法，并使用网格搜索优化超参数。这些优化算法的引入表明学者们开始关注特征选择和超参数调整对机器学习模型性能的影响，以及输入特征和超参数之间的复杂依赖性和相互作用，进而从模型层面提高预测精度。因此，之后的研究可以考虑引入不同优化算法，并结合特征变量筛选方法，以期进一步提高模型的反演精度和普适性。

本文选择光谱指数和雷达极化组合指数协同VIP-RF模型反演研究区耕地土壤含盐量，结果表明，研究区耕地盐渍化已较为严重，盐渍化土壤占耕地总面积的86.81%，中度盐渍化土壤分布范围最广，达到了耕地总面积的33.54%，这与前人研究[44]所得结论相符。平罗县地处内陆干旱区，气候干燥，土壤透气、透水性差，易产生盐渍化。中西部为山前洪积扇区和西大滩碟形洼地，排水条件差。东部引黄灌溉区灌溉方式不合理，常出现过量引水和引黄河大水漫灌现象，灌溉水的渗漏引起地下水位升高和强烈蒸发，造成盐渍化愈演愈烈，进而严重影响农业生产。因此，利用遥感监测，结合盐渍土改良与治理的主要机制，如物理调控、化学调理、灌排管理和生物改良[45]，因地制宜，提出可行性的治理措施，以减少盐渍化给当地农业发展和自然环境造成的危害。

本文土壤样本采样与影像获取时间不一致，而土壤含盐量存在短期变化情况，影像数据难以真实反应地表情况，数据存在不确定性。后续研究可以考虑利用无人机技术，在同一时间段采集土壤样本和遥感数据，且无人机数据分辨率高达厘米级，模型反演精度会更高。

4 结 论

本文采用VIP和GC筛选特征变量，利用光谱指数、雷达极化组合指数、光谱和雷达指数组合的3种不同数据源构建基于不同机器学习算法的土壤盐分反演模型，得到以下结论：

（1）基于GC方法确定的光谱指数和雷达极化组合指数数量差异较大，筛选变量所构建的模型反演效果较差；基于VIP方法确定的光谱指数和雷达极化组合指数数量相同，筛选变量所构建的模型精度相对较高。

（2）RF模型反演效果最优，SVM次之，BPNN效果最差，说明RF较其他2种机器学习算法更适于该研究区耕地土壤含盐量反演。

（3）组合多源遥感数据构建的VIP-RF模型验证集R^2=0.780，较单一光谱指数模型和雷达极化组合指数模型分别提高0.091和0.237，说明组合多源遥感数据能够有效提高土壤含盐量反演精度，有助于更加精确地研究区域土壤盐分分布情况，提升土壤盐渍化监测水平。

（4）根据反演结果可知，平罗县耕地盐渍化较为严重，非盐渍化土壤仅占耕地总面积的13.19%，轻度盐渍化和中度盐渍化土壤较多，占比分别为23.77%和33.54%，29.5%的耕地恶化为重度盐渍化和盐土。

参考文献（References）

[1] Abdelaziz M E, Abdelsattar M, Abdeldaym E A, et al. Piriformospora indica alters Na^+/K^+ homeostasis, antioxidant enzymes and LeNHX1 expression of greenhouse tomato grown under salt stress [J]. Scientia Horticulturae, 2019, 256: 108532, doi: 10.1016/j.scienta.2019.05.059.

[2] Singh A. Soil salinity: A global threat to sustainable development [J]. Soil Use and Management, 2022, 38(1): 39-67.

[3] 李星佑，张飞，王筝. 土壤盐渍化遥感监测模型构建方法现状与发展趋势[J]. 自然资源遥感, 2022, 34(4): 11-21. [Li Xingyou, Zhang Fei, Wang Zheng. Present situation and development trend in building remote sensing monitoring models of soil salinization [J]. Remote Sensing for Natural Resources, 2022, 34(4): 11-21.]

[4] 陈红艳，赵庚星，陈敬春，等. 基于改进植被指数的黄河口区盐渍土盐分遥感反演[J]. 农业工程学报, 2015, 31(5): 107-114. [Chen Hongyan, Zhao Gengxing, Chen Jingchun, et al. Remote sensing inversion of saline soil salinity based on modified vegetation index in estuary area of Yellow River[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(5): 107-114.]

[5] Wang J Z, Ding J L, Ma X K, et al. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China[J]. Geoderma, 2019, 353: 172-187.

[6] 赵孟银. 遥感影像去云方法研究[D]. 天津: 天津科技大学, 2016. [Zhao Mengyin. Study of clouds removal methods on remote sensing images[D]. Tianjin: Tianjin University of Science and Technology, 2016.]

[7] 李彪, 王耀强. 土壤盐渍化雷达反演模拟研究[J]. 干旱区资源与环境, 2015, 29(8): 180-184. [Li Biao, Wang Yaoqiang. Radar inversion and simulation of salty soil salinization[J]. Journal of Arid Land Resources and Environment, 2015, 29(8): 180-184.]

[8] 马驰. 基于Sentinel 1双极化雷达影像的土壤含盐量反演[J]. 农业工程学报, 2018, 34(2):153-158. [Ma Chi. Quantitative retrieval of soil salt content based on Sentinel-1 dual polarization radar image[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(2): 153-158.]

[9] 张智韬, 杜瑜燕, 劳聪聪, 等. 基于雷达遥感的不同深度土壤含盐量反演模型[J]. 农业机械学报, 2020, 51(10): 243-251. [Zhang Zhitao, Du Yuyan, Lao Congcong, et al. Inversion model of soil salt content in different depths based on radar remote sensing [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(10): 243-251.]

[10] 肖森天, 依力亚斯江·努尔麦麦提, 努尔比耶·穆合塔尔, 等. 基于光学和雷达多源遥感的于田绿洲土壤盐渍化时空分析[J]. 干旱区研究, 2023, 40(1): 59-68. [Xiao Sentian, Nurmemet Ilyas, Muhetaer Nuerbiye, et al. Spatial and temporal analysis of soil salinity in Yutian Oasis by combined optical and radar multi-source remote sensing[J]. Arid Zone Research, 2023, 40(1): 59-68.]

[11] 王飞, 丁建丽, 魏阳, 等. 基于Landsat系列数据的盐分指数和植被指数对土壤盐度变异性的响应分析——以新疆天山南北典型绿洲为例[J]. 生态学报, 2017, 37(15): 5007-5022. [Wang Fei, Ding Jianli, Wei Yang, et al. Sensitivity analysis of soil salinity and vegetation indices to detect soil salinity variation by using Landsat series images: Applications in different oases in Xinjiang, China[J]. Acta Ecologica Sinica, 2017, 37(15): 5007-5022.]

[12] 王海峰. 基于高光谱遥感的土壤水盐监测模型研究[D]. 咸阳: 西北农林科技大学, 2019. [Wang Haifeng. Hyperspectral remote sensing based models for soil moisture and salinity prediction[D]. Xianyang: Northwest A & F University, 2019.]

[13] 刘恩, 王军涛, 常步辉, 等. 小开河引黄灌区土壤盐渍化定量遥感反演[J]. 中国农村水利水电, 2019(12): 20-24. [Liu En, Wang Juntao, Chang Buhui, et al. Quantitative remote sensing inversion of soil salinization in Xiaokaihe Yellow River irrigation district[J]. China Rural and Hydropower, 2019(12): 20-24.]

[14] 刘亚秋. 黄河三角洲盐渍土盐分及组分定量光谱估测研究[D]. 泰安: 山东农业大学, 2017. [Liu Yaqiu. Quantitative estimation of soil salinity and its composition based on spectra in Yellow River Delta[D]. Tai'an: Shandong Agricultural University, 2017.]

[15] 曹肖奕, 丁建丽, 葛翔宇, 等. 基于不同卫星光谱模拟的土壤电导率估算研究[J]. 干旱区地理, 2020, 43(1): 172-181. [Cao Xiaoyi, Ding Jianli, Ge Xiangyu, et al. Estimation of soil conductivity based on spectral simulation of different satellites[J]. Arid Land Geography, 2020, 43(1): 172-181.]

[16] 魏慧敏, 贾科利, 张旭, 等. 基于机器学习和多光谱遥感的银川平原土壤盐分预测[J]. 干旱区地理, 2023, 46(1): 103-114. [Wei Huimin, Jia Keli, Zhang Xu, et al. Prediction of soil salinity based on machine learning and multispectral remote sensing in Yinchuan Plain[J]. Arid Land Geography, 2023, 46(1): 103-114.]

[17] 李小雨, 贾科利, 魏慧敏, 等. 基于随机森林算法的土壤含盐量预测[J]. 干旱区研究, 2023, 40(8): 1258-1267. [Li Xiaoyu, Jia Keli, Wei Huimin, et al. Prediction of soil salt content based on random forest algorithm[J]. Arid Zone Research, 2023, 40(8): 1258-1267.]

[18] Khan N M, Rastoskuev V V, Sato Y, et al. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators[J]. Agricultural Water Management, 2005, 77(1-3): 96-109.

[19] 樊彦国, 张维康, 刘敬一. 基于植被指数-盐分指数特征空间的黄河三角洲盐渍化遥感监测研究[J]. 山东农业科学, 2016, 48(5): 137-141. [Fan Yanguo, Zhang Weikang, Liu Jingyi. Remote sensing monitoring model of soil salinization in the Yellow River Delta zone based on vegetation index-salt index feature space[J]. Shandong Agricultural Sciences, 2016, 48(5): 137-141.]

[20] Sahana M, Rehman S, Patel P P, et al. Assessing the degree of soil salinity in the Indian Sundarban Biosphere Reserve using measured soil electrical conductivity and remote sensing data-derived salinity indices[J]. Arabian Journal of Geosciences, 2020, 13(24): 1289, doi: 10.1007/s12517-020-06310-w.

[21] Nguyen K A, Liou Y A, Tran H P, et al. Soil salinity assessment by using near-infrared channel and vegetation soil salinity index derived from Landsat 8 OLI data: A case study in the Tra Vinh Province, Mekong Delta, Vietnam[J]. Progress in Earth and Planetary Science, 2020, 7(1): 1-16.

[22] 孙亚楠, 李仙岳, 史海滨, 等. 基于多源数据融合的盐分遥感反演与季节差异性研究[J]. 农业机械学报, 2020, 51(6): 169-180. [Sun Ya'nan, Li Xianyue, Shi Haibin, et al. Remote sensing inversion of soil salinity and seasonal difference analysis based on multi-source data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6): 169-180.]

[23] Brunner P, Li H T, Kinzelbach W, et al. Generating soil electrical conductivity maps at regional level by integrating measurements on the ground and remote sensing data[J]. International Journal of Remote Sensing, 2007, 28(15-16): 3341-3361.

[24] 何宝忠, 丁建丽, 刘博华, 等. 渭库绿洲土壤盐渍化时空变化特征[J]. 林业科学, 2019, 55(9): 185-196. [He Baozhong, Ding Jianli, Liu Bohua, et al. Spatiotemporal variation of soil salinization in Weigan-Kuqa River delta oasis[J]. Scientia Silvae Sinicae, 2019, 55(9): 185-196.]

[25] 郭晓静. 内蒙古河套灌区冻融期土壤盐分的多极化雷达响应分析[D]. 呼和浩特: 内蒙古农业大学, 2014. [Guo Xiaojing. Response analysis of multi-polarization radar on freeze-thaw period

soil salt in Inner Mongolia Hetao irrigation district[D]. Hohhot: Inner Mongolia Agricultural University, 2014.]

[26] 于海云, 王志军, 李彪, 等. 内蒙古河套灌区融解期土壤盐分多极化雷达响应分析[J]. 长江科学院院报, 2015, 32(11): 19-24. [Yu Haiyun, Wang Zhijun, Li Biao, et al. Response analysis of multi-polarization radar on soil salt during melting period in Hetao irrigation district of Inner Mongolia[J]. Journal of Yangtze River Scientific Research Institute, 2015, 32(11): 19-24.]

[27] 李生勇, 霍轶珍, 王海霞. 内蒙古自治区河套灌区冻结期土壤盐分的多极化雷达响应分析[J]. 江苏农业科学, 2016, 44(3): 348-352. [Li Shengyong, Huo Yizhen, Wang Haixia. Response analysis of multi-polarization radaron soil salinity during freezing period in Hetao irrigation district of Inner Mongolia Autonomous Region[J]. Jiangsu Agricultural Sciences, 2016, 44(3): 348-352.]

[28] Oussama A, Elabadi F, Platikanov S, et al. Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores[J]. Journal of the American Oil Chemists Society, 2012, 89(10): 1807-1812.

[29] Maimaitiyiming M, Ghulam A, Bozzolo A, et al. Early detection of plant physiological responses to different levels of water stress using reflectance spectroscopy[J]. Remote Sensing, 2017, 9(7): 745, doi: 10.3390/rs9070745.

[30] 刘思峰, 蔡华, 杨英杰, 等. 灰色关联分析模型研究进展[J]. 系统工程理论与实践, 2013, 33(8): 2041-2046. [Liu Sifeng, Cai Hua, Yang Yingjie, et al. Advance in grey incidence analysis modelling [J]. Systems Engineering-Theory and Practice, 2013, 33(8): 2041-2046.]

[31] 杨淑娥, 黄礼. 基于BP神经网络的上市公司财务预警模型[J]. 系统工程理论与实践, 2005, 25(1): 12-18. [Yang Shu'e, Huang Li. Financial crisis warning model based on BP neural network[J]. Systems Engineering-Theory & Practice, 2005, 25(1): 12-18.]

[32] 赵亚琴. 基于模糊神经网络的火灾识别算法[J]. 计算机仿真, 2015, 32(2): 369-373. [Zhao Yaqin. Forest fire recognition algorithm based on fuzzy neural network[J]. Computer Simulation, 2015, 32(2): 369-373.]

[33] 汪海燕, 黎建辉, 杨风雷. 支持向量机理论及算法研究综述[J]. 计算机应用研究, 2014, 31(5): 1281-1286. [Wang Haiyan, Li Jianhui, Yang Fenglei. Overview of support vector machine analysis and algorithm[J]. Application Research of Computers, 2014, 31(5): 1281-1286.]

[34] Biau G, Scornet E. A random forest guided tour[J]. Test, 2016, 25 (2): 197-227.

[35] 杨练兵, 陈春波, 郑宏伟, 等. 基于优化随机森林回归模型的土壤盐渍化反演[J]. 地球信息科学学报, 2021, 23(9): 1662-1674. [Yang Lianbing, Chen Chunbo, Zheng Hongwei, et al. Retrieval of soil salinity content based on optimized random forest regression model[J]. Journal of Geo-information Science, 2021, 23(9): 1662-1674.]

[36] 余祥伟, 薛东剑. 星载SAR与OLI影像融合方法研究及质量评价[J]. 西华师范大学学报(自然科学版), 2020, 41(1): 98-103. [Yu Xiangwei, Xue Dongjian. Research and quality evaluation of spaceborne SAR and OLI image fusion methods[J]. Journal of China West Normal University (Natural Sciences Edition), 2020, 41 (1): 98-103.]

[37] 姜红, 玉素甫江·如素力, 拜合提尼沙·阿不都克日木, 等. 基于支持向量机回归算法的土壤水分光学与微波遥感协同反演[J]. 地理与地理信息科学, 2017, 33(6): 30-36. [Jiang Hong, Rusuli Yusufujiang, Abudoukerimu Baihetinisha, et al. Soil moisture retrieval by synergizing optical and microwave remote sensing data based on support vector machine regression algorithm[J]. Geography and Geo-information Science, 2017, 33(6): 30-36.]

[38] 李明亮. 土壤有机质高光谱灰色关联度估测模型研究[D]. 泰安: 山东农业大学, 2017. [Li Mingliang. Study on the grey relation degree estimation on model of soil organic matter based on hyperspectral[D]. Tai'an: Shandong Agricultural University, 2017.]

[39] Viscarra Roseel R A, Webster R. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database[J]. European Journal of Soil Science, 2012, 63(6): 848-860.

[40] 张智韬, 魏广飞, 姚志华, 等. 基于无人机多光谱遥感的土壤含盐量反演模型研究[J]. 农业机械学报, 2019, 50(12): 151-160. [Zhang Zhitao, Wei Guangfei, Yao Zhihua, et al. Soil salt inversion model based on UAV multispectral remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 151-160.]

[41] Chen B L, Zheng H W, Luo G P, et al. Adaptive estimation of multi-regional soil salinization using extreme gradient boosting with Bayesian TPE optimization[J]. International Journal of Remote Sensing, 2022, 43(3): 778-811.

[42] Wang L Y, Hu P, Zheng H W, et al. Integrative modeling of heterogeneous soil salinity using sparse ground samples and remote sensing images[J]. Geoderma, 2023, 430: 116321, doi: 10.1016/j.geoderma.2022.116321.

[43] Xu H T, Chen C B, Zheng H W, et al. AGA-SVR-based selection of feature subsets and optimization of parameter in regional soil salinization monitoring[J]. International Journal of Remote Sensing, 2020, 41(12): 4470-4495.

[44] 张旭. 银川平原土壤盐分空间分布及驱动因子研究[D]. 银川: 宁夏大学, 2022. [Zhang Xu. Study on spatial distribution and driving factors of soil salinity in Yinchuan Plain[D]. Yinchuan: Ningxia University, 2022.]

[45] 杨劲松, 姚荣江, 王相平, 等. 中国盐渍土研究: 历程, 现状与展望[J]. 土壤学报, 2022, 59(1): 10-27. [Yang Jinsong, Yao Rongjiang, Wang Xiangping, et al. Research on salt-affected soils in China: History, status quo and prospect[J]. Acta Pedologica Sinica, 2022, 59(1): 10-27.]

Inversion of soil salt content by combining optical and microwave remote sensing in cultivated land

LIU Ruiliang[1], JIA Keli[1], LI Xiaoyu[1], CHEN Ruihua[1],
WANG Yijing[1], ZHANG Junhua[2]
(1. College of Geographical Sciences and Planning, Ningxia University, Yinchuan 750021, Ningxia, China;
2. College of Ecology, Ningxia University, Yinchuan 750021, Ningxia, China)

Abstract: The safeguarding of cultivated land is paramount in ensuring national food security, sustainable economic and social development, and the preservation of the ecological environment. Rapid and accurate acquisition of cultivated soil salinity and spatial distribution information is imperative for the protection of cultivated land. This study focuses on cultivated land in Pingluo County, Ningxia, China, as the research object and discusses the feasibility of combining optical remote sensing and microwave remote sensing to predict the accuracy of soil salt content compared with single remote sensing data. The methodology involved the extraction of the spectral indices from Landsat 9 OLI and radar polarization combination indices from Sentinel−1. Variable projection importance and gray correlation degree were used to screen and combine characteristic variables. Three machine learning algorithms (back propagation neural network, support vector machine, and random forest) were used to construct the soil salt content prediction model. The best model was used to predict the spatial distribution of the soil salt content in cultivated land. The results show the following facts: (1) The model, validated using the variable projection importance method for screening variables, generally exhibited a higher determination coefficient (R^2) than the model established using the gray correlation method for characteristic variables. (2) Using the random forest algorithm, the model combining the spectral index and radar polarization combination index demonstrated the best effect. The modeling set exhibited an R^2 of 0.791 and a root mean square error (RMSE) of 1.016. This represented an increase in R^2 by 0.065 and 0.085 compared with the single data source model, with corresponding decreases in RMSE by 0.147 and 0.189. The validation set showed an R^2 of 0.780 and an RMSE of 1.132, indicating a respective increase in R^2 by 0.091 and 0.237 and a decrease in RMSE by 0.175 and 0.377 compared with the single data source model. (3) The distribution range of mildly salinized and moderately salinized soil of cultivated land in Pingluo County covered wide areas, accounting for 23.77% and 33.54%, respectively, whereas severely salinized soil constituted 15.37%. This underscores the effectiveness of modeling by combining multisource remote sensing data in improving the prediction accuracy of soil salt content. The outcomes offer a valuable technical reference for predicting soil salt content in arid areas and contribute to the sustainable development of local agriculture.

Key words: optical and microwave remote sensing; machine learning; cultivated land; inversion of soil salt content